

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)



10 030885



INVESTOR IN PEOPLE

GB00
01610

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

ESU

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

REC'D 18 AUG 2000

WIPO

FCI

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

M. C. E. K.

Dated

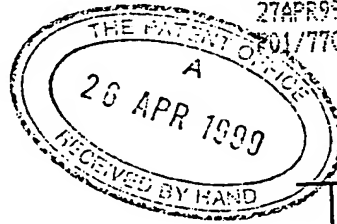
2.8.00

THIS PAGE BLANK (USPTO)

For Official use only

9909606.7

Your reference Download Parameters (C)



27APR99 E442900-2 003274
901/7700 0.00 - 9909606.7

The
**Patent
Office**

Request for grant of a
Patent

Form 1/77

Patents Act 1977

1 Title of invention

Networked delivery of profiled media files to clients

2. Applicant's details



First or only applicant

2a

If applying as a corporate body: Corporate Name
Telemedia Systems Limited

Country

Great Britain

2b

If applying as an individual or partnership
Surname

Forenames

2c

Address

Mount Pleasant House
2 Mount Pleasant
Huntingdon Road
Cambridge

UK Postcode

CB3 0RN

Country

Great Britain

ADP Number

07647860001.

☐

Second applicant (if any)

2d

Corporate Name

Country

2e

Surname

Forenames

2f

Address

UK Postcode

Country

ADP Number

3 Address for service

Agent's Name

Origin Limited

Agent's Address

24 Kings Avenue
London

Agent's postcode

N10 1 PB

05627856002.

Agent's ADP
Number

C03274

4 Reference Number

Download Paramters (C)

5 Claiming an earlier application date

An earlier filing date is claimed:

Yes ☐

No ☒

Number of earlier
application or patent number

Filing date

15 (4) (Divisional)

☐

8(3)

☐

12(6)

☐

37(4)

☐

6 Declaration of priority

Country of filing

Priority Application Number

Filing Date

--	--	--

7 Inventorship

The applicant(s) are the sole inventors/joint inventors

Yes ☐

No ☒

8 Checklist

Claims

23 /

Continuation sheets

1 / 23

Description

22

Abstract

1 /

Drawings

9 + 9

Priority Documents

Translations of Priority Documents

Patents Form 7/77

Patents Form 9/77

Patents Form 10/77

9 Request

We request the grant of a patent on the basis of this application

Signed:

Origin

(Origin Limited)

Date:

24 April '98

Networked delivery of profiled media files to clients

Technical Field

5 This invention relates to methods for the delivery of profiled media files, such as video and audio files, from a server to a client using a network, in which the files stored on the server are encoded so that the generation and/or reception of their associated bitstreams can be encoded to yield a media file at a client which meets specific criteria, such as pre-defined image quality and spatial resolution levels.

10

Background to the Invention

Recent years have seen considerable interest in improving techniques for delivering all kinds of media files, such as video and audio files, to PC clients over low bandwidth networks, prompted not least by the ever-increasing use of still and animated images in Web pages. In all prior art cases, the quality of the resulting images at the client and the amount of data transferred from the server to a client is however not only determined by the server that is delivering the data but is also the same for all clients.

20 Hence, no account is taken of the fact that different clients will likely have different connection bandwidths and may wish to see an image at different quality levels. For example, a high-quality colour image may be delivered by the server only to be displayed at low resolution in monochrome by the client. Alternatively, the result can be poor quality images despite the availability of a significant bandwidth from the server. For example, a Web site may use low-quality GIF images to support low-bandwidth modem users that are inappropriate for clients with high-bandwidth ADSL connections. The problems are increased when animated sequences are used because of the increased amount of data that must be delivered. Generally, media file delivery over reduced bandwidth connections such as the Internet is currently characterised by the delivery of unwanted data and significant delays in the display of information.

30

One approach to addressing these problems is found in the Custom Download Configuration option in the Indeo product from Intel Corporation. In this system, one encoded video file can be played back at many different data rates and quality levels, each of which is suitable for different network bandwidths or playback-platform capabilities. But

35

the file author has to set the relevant frame rate and the required quality level at a server prior to a download, so that the download meets these quite limited author determined requirements. Once set, the download configuration will apply to all downloads to all clients. Reference may also be made to US patent no. US5585852 to Intel Corporation
5 which discloses a system in which a source video is transformed into 6 mini-videos. One or more of these 'band sequences' can be omitted in the inverse transform, thereby losing the information carried in the omitted band sequence. This yields a form of quality scalability, although it is inflexible.

10 Another problem when image browsing over a network is that an image is often viewable only after a significant portion of the download is complete; this can be both time consuming, slowing down the browsing of sites and, when the image proves unsuitable or uninteresting, frustrating. It is therefore very convenient if an image can become progressively clearer as the image download continues: a user therefore initially, and rapidly,
15 sees a very low resolution image, which gradually becomes clearer as more data is downloaded, allowing the user to end the download at any time. One conventional approach to meeting this need for 'progressive image transmission' is the interleaved GIF image. Another approach is disclosed in US Patent No. 5880856 to Microsoft Corporation, which teaches a particular approach to the use of wavelet transforms.

20

In the US 5880856 patent, an image is transformed using the wavelet transformation to yield 4 or 5 decomposition levels, with a base decomposition level giving a low resolution image, and increasingly higher decomposition levels giving higher resolution images. The client receives initially only the base decomposition level data, but the low resolution image
25 resulting from the base decomposition level data gradually sharpens up as higher decomposition levels are received and decoded. Sharpening up of the image occurs as a result of 2 factors: first, as all of the 4 sub-bands which form each decomposition level are received and decoded, the image quality level increases slightly. As successive decomposition levels are received and decoded, the image quality increases more
30 significantly. One characteristic of the system taught in the US 5880856 patent is that the sub-bands and decomposition levels are all transmitted to clients in the same, fixed order, which is relatively inflexible. As with the Indeo system from Intel Corporation, the server pre-determines the quality of the resulting image and the way in which the download progresses, applying the same criteria for all downloads to all clients.

35

Further reference may be made to Beong-Jo Kim, Zixiang Xiong and William Pearlman, "Very Low Bit-Rate Embedded Coding with 3D Set Partitioning in Hierarchical Trees", submitted to IEEE Trans. Circuits & Systems for Video Technology, Special Issue on Image and Video Processing for Emerging Interactive Multimedia Services, Sept. 1998.

5 This paper discloses applying a SPIHT compression scheme to a wavelet-based transform, which yields a bitstream encoding multiple spatial resolutions, with progressive quality ordering within a given spatial resolution, as in the US 5880856 patent. A particular feature is the use of 'flags' inserted by the client decoder into the received bitstream to mark temporal/spatial locations defined by the input resolution parameters. But because the
10 'flags' are inserted at the decoder, much (and in some cases, all) of the bitstream has to be received and stored at the client, so that considerable bandwidth may be wasted.

One final aspect of conventional systems is that, in the case where an image or sequences is used in a number of locations, caching mechanisms can be used to avoid repeated download
15 of the same data. However, if an image is required at a different resolution or quality, the entire data for the new image must conventionally be downloaded even if a higher quality version is already cached on the client.

20 Summary of the Invention

In accordance with the present invention, there is provided a method of delivering a media file to a client on a network in which a server derives the media file from a source file on the server and delivers the media file to the client, the derivation and/or delivery being in compliance with a set of download parameters, the download parameters defining one or
25 more download and/or media variables.

Hence, the present invention is predicated on the insight that it is possible to construct a system with a set of download parameters which enables the server and/or client to set useful download and/or image variables appropriate for a client, so that any client can
30 download and view the media file in the optimal manner for its own needs. This is particularly attractive in the context of browsing Web-based media files, such as images, audio and animations and is clearly more flexible than the conventional systems which restrict all clients to the fixed download parameters, uniform for all clients, which are set at a server. The term 'network' should be expansively construed to cover any kind of data
35 connection between 2 or more devices. A 'file' is any consistent set of data, so that a

'media file' is a consistent set of data representing one or more media samples, such as frames of video or audio levels.

- 5 Preferably the download parameters are either stored on the server or are transmitted to the server by the client before or during the delivery of the media file, or a combination of the two. Download parameters for a client are typically specific to the needs and circumstances of that client.
- 10 The download parameters may define one or more of the following media parameters for an image or sequence of images:
 - (i) the spatial resolution of the image or images;
 - (ii) the quality of the image or images;
 - (iii) the number of displayable frames;
 - 15 (iv) the preferred order in which data is to be transmitted;
 - (v) the selection of colour components within one or more frames;
 - (vi) a sub-set of frames to be delivered; and
 - (vii) a sub-region within one or more frames;
- 20 For audio, the download parameters may define one or more of the following media parameters:
 - (i) the quality of the audio;
 - (ii) the preferred order in which data is to be transmitted;
 - (iii) the number of audio channels to be transmitted; and
 - 25 (iv) selection of monophonic, stereophonic or quadraphonic audio.

Other parameters may include the following:

- (i) the rate at which data is to be transmitted to the client;
- (ii) the preferred order in which data is to be transmitted;
- 30 (iii) the set of data for the source file that is already stored on the client and does not need to be transmitted; and
- (iv) the maximum amount of data to be transmitted to the client;

The derivation of the media file may be adapted to take account of one or more of the following:

- (i) the data size of the original of the source file
- (ii) the bandwidth available to a client.
- (iii) the current or predicted loading of the server

5

Hence, the invention may provide reduced download times for clients requiring lesser quality, enhanced quality for higher-bandwidth users and a progressive image display that satisfies differing client quality requirements. For example, when browsing Web sites for information content rather than images, any images on a site can be download only at low quality initially, so that rapid browsing is not hampered by slow image downloads of marginal interest to a user. Additionally, any one or more of the download parameters may be altered during the transfer of the media file. Further, any one or more of the download parameters may be altered by the client after the transfer of the media file has completed; only the required extra data (if any) is then transferred from the server.

10

15

The client may also specify the order of progressive transmission. For example an animated sequence may be transmitted such that a low quality version of each frame is made available and played while the high quality data is transmitted. Alternatively the same file may be transmitted such that each frame in the sequence is transmitted at the required quality before the next frame is transmitted. In addition, the client may specify the maximum rate at which the data is delivered. When a number of images are required this allows the client to simultaneously receive all the images progressively, rather than waiting for one image to load before a transmission starts on the other images.

20

25

30

In one embodiment, the media file may be retained at the client and subsequent display of the file at a different resolution, quality or frame rate will then require only limited, additional data to be transferred to the client. Intelligent caching of the data is used to achieve this and allow the image to be re-displayed in an efficient manner even if the client requests different resolution or quality.

The present invention requires that the encoded data from the server has certain special properties, but is independent of any specific encoding scheme. However it works particularly well with wavelet-based schemes.

35

Hence, in a preferred embodiment, the media file is generated using a wavelet transform, the output of which may be compressed using SPIHT or other forms of compression. The wavelet approach is a particularly powerful technique for providing several useful features: image quality may progressively increase as the sequence of media files is transmitted and the number of displayable frames in a sequence may progressively increase as the sequence of media files is transmitted. A download can also be halted at any time resulting in a displayable image or sequence of images, albeit at reduced quality from the final version.

In another preferred embodiment, the data in the media file is structured by an encoder as a bitstream including several discrete bitstream layers, in which layer signalling information which identifies individual layers is inserted by the encoder, the layer signalling information enabling each client to be sent only those layers which satisfy the download parameters specified by that client. A media server may store the media files, and can distribute to different networked clients bitstreams with different properties depending upon the layers which satisfy the download parameters associated with each client. The terms 'layer' and 'layer signalling information' are defined and expanded upon in a later section of this specification.

Other aspects of the invention relate to a media file which is deliverable using any of the inventive methods defined above; a computer program which when running on a client enables the client to receive and playback a media file delivered using any of the above methods; and a computer program which when running on a server or encoder enables the server or encoder to perform any of the above methods.

In a final aspect, there is provided a server programmed to deliver to a client on a network a media file deriving from a source file stored on the server, wherein the server is programmed to derive the media file from a source file on the server and deliver the media file to the client, the derivation and/or delivery being in compliance with a set of download parameters, the download parameters defining one or more download and/or media variables.

Brief Description of the Drawings

The invention will be described with reference to the accompanying drawings, in which:

Figure 1 is a schematic representing the sub-bands which result from applying wavelet transforms to an image in a conventional multi-scale decomposition;

5 Figure 2 is a schematic representing the sub-bands which can be re-constructed in a multi-scale reconstruction;

Figure 3 is a schematic representation of a network used in performing the method of the present invention;

10 Figure 4 is a schematic representation of an encoder used in performing the method of the present invention;

15 Figure 5 is a schematic representation of a media server used in performing the method of the present invention;

Figure 6 is a schematic representing a hierarchical decomposition of a particular media object (a video clip) into a set of layers;

20 Figures 7 to 11 are schematics representing various hierarchical decompositions of a particular media object (a video clip) into a set of layers.

Detailed Description

25

An Overview

Image Encoding

30 The method of the present invention requires, in one embodiment, that image data be encoded in such a way that the resulting data for a frame or sequence of frames can be partitioned into two or more distinct sections using an encoding process. Different sections of the data will encode different parts of the image or sequence of images at different resolutions and quality. It must be possible to decode a sub-set of the complete data set in order to re-construct the frame or sequence of frames at a particular displayed resolution, quality and frame rate.

35

By way of preferred example, partitioning of a bitstream generated using a wavelet transform and SPIHT compression into 'layers' (as that term is defined in this specification) is described later on in this specification. 'Layers' are one example of what a section might be in an embodiment. 'Layers' exemplify the concept of a section as defined above since they allow encoding of (inter alia) different spatial resolutions (or scales) and quality.

By decoding one or more of these sections or 'layers' of data it is possible to re-construct the frame or sequence of frames at a particular displayed resolution, quality and frame rate. In general, the quality of the resulting frame or sequence of frames will depend on the total amount of data that is being decoded. The choice of sections (or 'layers' in the wavelet/SPIHT example) that are used to re-construct the frame or sequence of frames can be made on the basis of the desired resolution, quality and frame rate. In the wavelet/SPIHT example given later on, choice is exercised using a 'Client Layer Template'. Also, it is possible to improve the quality, resolution and frame rate of an existing frame or sequence of frames by adding more data sections. This process is known as *enhancement*. (In the wavelet-based example given later, we use the term *refinement* for this process.)

Labelling of Image Sections

The method of the present invention involves, in one embodiment, labelling each data section with a number of pieces of information to allow the server to deliver the data according to the client requirements. This labelling may be explicit or implicit in the data format. The information must include the following:

- ◆ The size of the data section
- ◆ A unique identifier for the data section
- ◆ A list of data sections that must be present for the section to be decoded, known as the *parent sections*
- ◆ A value indicating the quality of image that would result from decoding this and all its parent sections
- ◆ A value indicating the scale (normal displayed resolution) of the image that would result from decoding this and all its parent sections
- ◆ The frame or range of frames that are include in the section

It may also include one or more of the following:

- ◆ The region or regions of the frame or frames that are included in the section

- ◆ The colour components that are contained within the section.

In addition, the server maintains a set of information relating to the whole file. This includes the following, where appropriate:

- 5 ◆ The number of frames in the file
- ◆ The display time of each frame (for example, for animation purposes)

Some aspects of section labelling are achieved in the wavelet implementation described later in this specification using the technique referred to as 'layer signalling'.

10

Client-driven image delivery

The method described in this specification allows a client to specify information about an image that it wishes to receive and have the server deliver the appropriate sections of the image data at the appropriate time. The client can specify (amongst other parameters) the image size, the image quality, the frame-rate, the data rate and the total amount of data that is required. The server will use this information to decide which sections of the image data should be sent to the client and at what rate. In the wavelet/SPIHT example given later, this process is described as 'layer multiplexing'.

- 20 The Client will set the parameters on the basis of user requirements and the current state of the client. For example, the image quality may be explicitly set by the user; the displayed resolution may be determined by the size of the display window that has been selected; the set of sections to be omitted may be determined by the sections that are available via an alternative mechanism, such as (but not restricted to) a data cache in memory or on disc. In
- 25 the example of an image embedded in a Web Page, some parameters may be specified within the web page itself or in an associated page.

Download Parameters

When a client requests the delivery of data it specifies a set of *download parameters* that may include any of the following:

- 30 1) Parameters that specify the required scale (or spatial resolution) of the final sequence of images
- 2) Parameters that specify the required quality of the final sequence of images
- 3) Parameters that specify the required frame-rate of a sequence of images
- 35 4) Parameters that specify the colour requirements

- 5) Parameters that specify the frame or set of frames that are required
- 6) Parameters that specify the region or regions of the image that are required
- 7) Parameters that specify the rate at which the data is to be delivered
- 8) Parameters that specify the progressive order in which the data is to be transmitted
- 5 9) Parameters that specify the set of data that is already held by the client and therefore does not need to be transmitted.
- 10) The maximum amount of data to be transmitted

10 In the wavelet/SPIHT example, the download parameters are contained in the Client Layer Template.

Server selection of data sections

15 The server uses the download parameters to deliver the appropriate data to the client; this is the 'layer multiplexing' process described subsequently in relation to a wavelet embodiment of this invention. This process involves the following steps:

- 1) The server determines the total set of data sections to be transmitted
- 2) The server determines the order in which the data sections are to be transmitted
- 3) The server delivers the sections in order, at the specified rate

20 Determining the set of data sections

The server selects the data sections that need to be transmitted on the basis of some of the download parameters. It starts with a complete set of data sections and then discards sections on the following basis:

- 1) It discards all sections that code colour components that are not required.
 - 25 2) It discards all sections that code only for frames that lie outside the specified set of frames.
 - 3) It discards all sections that code only for frames that are not required for the selected frame rate.
 - 4) It discards all sections that code for a quality greater than the required quality that have parent sections that code for at least the required quality.
 - 30 5) It discards all sections that code for a scale greater than the required scale that have parent sections that code for at least the required scale.
 - 6) It discards all sections that are marked as already available on the client.
- Other selection criteria may also be applied at this point.

Ordering the data sections

Having selected the required sections, the server then orders the data sections for delivery. The data sections are sorted according to a list of criteria. The most important criterion is the dependence between data sections and their parents. The parent of a data section always appears before the data section in the list to ensure that every data section can be decoded as soon as it arrives.

The remaining criteria and their relative priority are specified by the client in the download parameters. The following criteria can be applied:

- 1) The data sections can be compared by the scale for which they code.
- 2) The data sections can be compared by the quality for which they code.
- 3) The data sections can be compared by the first frame number for which they code.
- 4) The data sections can be compared by the region for which they code
- 5) The data sections can be compared by the colour components that they contain.

Other ordering criteria may also be applied at this point.

Delivery of the data sections

The server delivers the selected sections in the selected order to the client at the rate specified in the download parameters.

Client handling of received data

When at least one complete data section has been received by the client it may decode and display the image or images encoded by those data sections. The data sections are presented to the decoder which then generates the decompressed data. The client may elect to do this on a single occasion when all the requested data has been delivered. Alternatively it may elect to decode the data when only a partial section of the data has been delivered, in order to implement the progressive download feature. A data section cannot be fully decoded unless the specified parent sections are also available. Therefore a client may elect to decode only the sub-set of the available data sections for which this condition applies.

In the case where the data connection from the server to the client is unreliable, it is possible that some of the requested data sections may not arrive, or may arrive in a corrupt form. In

this case the client may change the download parameters to request the re-transmission of the relevant sections.

Changes to download parameters

- 5 The client may change the download parameters at any time. The server will re-calculate the new set of required data sections, allowing for the sections that have already been delivered to the client. If a section is currently being transmitted, transmission will be completed unless the section no longer appears in the required list, in which case it will be terminated. If the transmission is terminated the data stream marked accordingly to notify the client of
- 10 the incomplete data section.

Extension to other encoding means of derivation

- The derivation of the media file from the source file may involve further processing of the source file. For example, a source file may consist of a set of co-ordinates for lines and text
- 15 in a diagram. In order to derive the media file from this source file the server may render the lines and text at the requested resolution and then encode the resulting image in a compressed format such as GIF.

Extension of the invention to other media types

- 20 The application of the invention as presented is not limited to media files representing images. It may be applied to other static or continuous media formats. The extension of the invention to cover a new medium requires changes to the set of supported download parameters. This allows the data sections to be selected on the basis of new criteria. The underlying mechanism of the invention is unaffected. In particular, the mechanism may be
- 25 applied to audio files encoded in an appropriate way.

Extension of the invention to audio

- The extension of the invention to cover audio requires changes and additions to the set of download parameters. The parameters that relate to image-specific features, such as colour,
- 30 are no longer required. Other parameters can be adapted to the new medium. For example, the quality parameter can be used to specify the quality of the sound that is heard rather than the quality of the displayed image. New parameters can be introduced to cover audio-specific features. For example, the stored original may contain multiple tracks for stereophonic playback. The client can specify that only a single channel, or a monophonic
- 35 version of both channels, is delivered.

Benefits

Reduced storage

Because of the nature of the encoding for the image data it is possible to meet a number of different requests from the same file. This reduces the storage requirement on the server machine. For example, a Web page may contain a thumbnail version of a photograph as well as a full-sized version. With existing technology this is usually achieved by having two distinct versions of the image on the server. Using the method described in this specification, a single version of the image can be used in both places.

Reduced network bandwidth

The nature of the image encoding can also reduce the amount of data that needs to be sent over the network, thus reducing download times. The server will only send the data that is required to display the image at the selected resolution. This is in contrast to existing solutions where images are sometimes shrunk after they are downloaded in order to fit into a smaller space. This leads to unneeded data being downloaded and slower display of web pages.

Improved caching

The client can retain the information for an image that was downloaded at a particular resolution and quality and re-use it when the image is re-displayed. If the image is re-displayed at a lower resolution and quality, the existing data can be used and no new data needs to be downloaded. If the image is re-displayed at a higher quality or resolution, the existing data can be enhanced and only the extra data sections need to be downloaded.

A Wavelet Implementation

Some key concepts

The wavelet transform has been intensely studied for many years. Reference may for example be made to Mallat, Stephane G. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.11, No.7, pp 674-692 (Jul 1989). The function of the wavelet transform is to compact most of the energy in an image to a small number of pixels. The transform also produces a set of sub-bands consisting mainly of zero or low-value coefficients; it therefore generates a large number of pixels with small or zero values.

Because of the large number of pixels with small or zero values, there is considerable scope for a compression.

SPIHT compression is under increasing scrutiny: see for example Beong-Jo Kim, Zixiang Xiong and William Pearlman, "Very Low Bit-Rate Embedded Coding with 3D Set Partitioning in Hierarchical Trees", submitted to IEEE Trans. Circuits & Systems for Video Technology, Special Issue on Image and Video Processing for Emerging Interactive Multimedia Services, Sept. 1998. The SPIHT compression algorithm is particularly effective operating in conjunction with a wavelet transform, since it can readily exploit the wavelet transform output to get high levels of data reduction. Other compression schemes can also be used in the present invention, as will be appreciated by the skilled implementer.

The preferred wavelet transform in an embodiment of the present invention is Mallat's Fast Wavelet Transform (FWT), which generates a hierarchy of power-of-two images: it "critically" samples the image at each stage of the transform. Every time that it generates a 2^n image, the spatial sampling frequency - the 'fineness' of detail which is represented - is reduced by a factor of two in x and y. To do this, the wavelet filter removes all signal energy above (sample freq/2), and distributes it into the high-frequency sub-bands. These are the 'detail' images which compress very well because they are mostly zeroes or low-values. The smaller, critically-sampled (or smoothed) image which is generated at each stage can be viewed directly at that resolution because the high-frequencies which would otherwise cause aliasing have been removed. (Aliasing occurs when you try to render an image which contains frequency components too high for the display device. For example consider an image of a wooden fence where the uprights happen to lie on every other pixel. If you simply reduce the resolution by two by missing out every other pixel the fence is rendered either as solid black or it disappears.)

Scale

FWT employs the concept of 'scale'. A function can be described by analysing it into a set of discrete spatial/frequency components of differing *scales*. In the preferred embodiment of the present invention, FWT is combined with SPIHT such that the image scale ordering resulting from the transform is maintained through the property of significance ordering of SPIHT.

The operation of FWT on an image can be seen in Figure 1. Here, the original of an image is transformed or decomposed into 4 sub-bands at Level 0. Each sub-band is a quarter the size of the Original Image and a quarter the spatial resolution.

5 FWT operates as follows on a 2-D image:

- First, a high and low pass wavelet filter (i.e. a wavelet filter and a scaling function) are applied to each vertical line of pixels in the original, generating 2 sub-bands, one with high frequency information and the other with low frequency information.
- 10 • The filters are then applied to each horizontal line of pixels in the 2 sub-bands, splitting each sub-bands into two, yielding a total of 4 sub-bands, LL, LH, HL and HH. This is shown as Scale Level 0. LL contains low frequency information from the vertical and horizontal convolutions. If viewed, it would look like the Original, but at a quarter resolution and a quarter size. LH contains
- 15 low frequency information from the horizontal and high frequency from the vertical. HL contains high frequency from the horizontal and low frequency from the vertical. HH contains high frequency from the vertical and horizontal convolutions.
- A further decomposition can be performed on LL to yield a Scale Level 1 decomposition. In practice 4 to 7 successive decompositions may be applied.
- 20

Multi-Scale Reconstruction

Each subband describes the original image at a particular scale, with a particular sampling frequency. To reconstruct an image, inverse wavelet transforms are applied to the 4 sub-

25 bands shown at Scale level 1 to re-generate sub-band LL at Level 0. Combining re-generated LL with the other 3 sub-bands in Level 0 enables the Original image to be obtained. As will be explained later, a *Layer Multiplexing* process is free to combine any sub-bands to produce a final image most appropriate to the particular requirements of the client. *Layer Truncation* can also optionally be used to limit the bit-rate at the cost of

30 increased distortion.

Three examples are shown in Figure 2 which show the flexibility of the *Layer Multiplexing* approach. In the first example, Figure 2A, the LL sub-band of Level 1 is used: this will reconstruct a coarse version of the original, at one sixteenth the resolution and size of the

35 original. In the second illustration, Figure 2B, the LL sub-band of Level 0 is used: this will

reconstruct a larger and better quality image; it is one quarter the size and resolution of the original. In the third example, Figure 2C, the LL and LH sub-bands of Level 1 are used, together with the LH sub-band of Level 0. This reconstructs an original resolution image, but with horizontal features at all scales emphasised

5

System description

The system capable of performing the method of the present invention comprises a media encoder, network, media server and N clients, as shown in Figure 3. A media encoder processes the incoming media into a source file with a layered structure. The layering within the embedded bitstream may include any layering needed to support the differing download parameters which a client might specify to the server, such as *Region Layering* and *Significance/Scale Layering*. The encoder also inserts *Layer Signalling* prior to the source file being transmitted as a bitstream over a digital communications network, where it is stored on a media server.

15

In Figure 3, there are N clients, each of which engages in a browse session with the media server during which media content and control information are transmitted. For each client a control channel to the server is maintained which allows that client to request that media be transmitted to it. The requests that can be made include (but are not limited to) the following:-

20

- Seek to a particular point in the media file.
- Transmit a media clip from the seek point at a specified bit-rate using the *Client Layer Template* to determine exactly what parts of the media are to be sent.
- Transmit a *Refinement Layer* for a previously sent media clip.
- Update or configure the *Client Layer Template* for this client.

25

These requests support a variety of functions in the client. For example, in a browsing application, a *Base Layer* may initially be sent to the client. This allows rapid scanning of low-quality material using minimum system resources (channel capacity, decoder CPU, etc.). To play a section with greater quality, the client updates the *Client Layer Template* to specify the exact refinement information which should now be delivered to improve the media information already stored at the client. The embedded nature of the data means that no processing is required at the client other than simply to insert the new data into the correct positions in the stored bitstream. This refinement information may perform (but is not limited to) the following actions:-

35

- Increase the temporal resolution of the media (i.e., for video, allow more individual frames to be resolved).
- Increase the spatial resolution of the video.
- Increase the sampling frequency of the audio.
- Decrease the degree of distortion of the media.
- Improve the quality of a 2-dimensional area within the image.
- Improve the media content at particular scales (for example enhance fine horizontal details in the video, or emphasise high frequencies in audio).

10 All these actions can be used to support application-level tasks such as video previewing using fast-forward and rewind, single-frame stepping forwards and backwards, and freeze-frame. As explained earlier, this refinement (or enhancement) process can use data transferred using any surplus bandwidth arising in the network.

Description of the Encoder

15 As depicted in Figure 4, a media capture engine samples and stores incoming media at the source resolution. This is passed to a wavelet transform engine which performs *Multi-Scale Decomposition* of the source material into a wavelet coefficient representation.

20 A SPIHT encoder compresses the wavelet coefficients into an embedded bitstream. The bitstream encodes the original picture to a fidelity as near lossless as possible, subject to processing and network constraints, i.e., the encoding attempts to deliver A *Layer Signalling* engine inserts information which delimits the individual blocks of data in the significance/scale layered bitstream. The signalling allows any subband and any refinement level to be efficiently accessed by a *Layer Multiplexing* process on the media server.

25

Description of Layer Signalling

Each Layer is labelled with a number of pieces of information to allow the server to deliver the data according to the client requirements. The information may include, but is not limited to, the following:

- 30 (a) The size of the Layer
- (b) The type of Layer
- (c) A unique identifier or sequence number for the Layer data.
- (d) If the Layer is a spatial or temporal Scale Layer, as defined below, a value indicating the spatial or temporal resolution of the original media object from which this scale Layer is
- 35 derived.

- (e) If the Layer is a spatial or temporal Scale Layer derived from a wavelet transform, as defined below, a value indicating the level and subband for this scale.
- (f) If the Layer is a spatial Region Layer, as defined below, a value indicating the size, shape and position of the region.
- 5 (g) If the Layer is a temporal Region Layer, as defined below, a value indicating the frame or range of frames included in the region.
- (h) If the layer is a colour component Layer, as defined below, a value indicating the particular component.

10 **Description of Client Layer Template**

When a client requests the delivery of data it specifies a Client Layer Template that may include any of the following:

- (a) Parameters that specify which Layers are to be compiled into the bitstream.
- (b) Parameters that specify the rate at which the data is to be delivered
- 15 (c) Parameters that specify the progressive order in which the data is to be transmitted
- (d) Parameters that specify the set of data that is already held by the client and therefore does not need to be transmitted.
- (e) The maximum amount of data to be transmitted

20

Description of the Media Server

The embedded bitstream is stored as a file on the Media Server. In detail, this process involves the following steps, in which the Server:

- (a) locates the Client Layer Template for the client,
- 25 (b) determines the total set of Layers to be transmitted,
- (c) determines the order in which the Layers are to be transmitted,
- (d) locates Layer data which may be in fast temporary storage, such as cache memory, as a result of delivery to another client,
- (e) locates the remaining Layer data using file pointers derived from the index.data,
- 30 (f) delivers the Layer data in order, at the specified rate,
- (g) terminates the transmission when the specified maximum amount of data has been reached.

35 An indexing process scans the bitstream as it is input to produce a cross-reference file which relates layer data to positions in the file. This is used by the Layer Multiplexer

efficiently to create bitstreams with different properties for different clients. When a client requests media data the Layer Multiplexer first of all consults the Client Layer Template for that individual client to discover which layers of data are required, and at what bit-rate. It then consults the Index Data to retrieve file pointers for the required layers. The requested data is then transmitted to the client over the digital communications network.

Further Layer Multiplexing examples

Figure 6 shows an example of a hierarchical decomposition of a particular media object (a video clip) into a set of layers. The video is first of all transformed into one or more temporal regions (T-regions), where each T-region can encode different numbers of frames, with different encoding parameters, and different underlying layer structures. The ability to generate T-regions with different properties allows the encoder to respond to external stimuli and to adapt to new situations. For example in a surveillance application a signal from a motion detector can start the generation of a new T-region with a greater amount of information and resolving power in the underlying layers.

T-regions may be composed of one or more temporal scale (T-scale) layers, where a Scale Layer is a layer for which all the refinement information is for a particular Scale. Thus each T-scale captures a particular aspect of the temporal activity in the T-region. Each T-scale may be composed of one or more spatial regions (S-regions) for which all the underlying information is for a particular 2-dimensional area in the image plane. This enables different areas to be encoded with differing information content; in particular it allows 'hi-fidelity' windows to be defined over areas of particular interest.

Each S-region may be composed of one or more spatial scale (S-scale) layers, where each S-scale captures a particular aspect of the spatial activity in the S-region. Each S-scale may be composed of one or more component layers, where all the underlying information refers to a particular component in a colour system, for example, R,G,B or Y,U,V or other format.

Finally, the component layer is composed of one or more significance layers, as follows:

In an embedded, bit-plane oriented compression scheme such as SPIHT, compression is obtained by partially ordering all the samples with respect to bit-significance, i.e., the position at which the highest-order bit in a sample is set. Since this corresponds to the magnitude of the sample, the samples are effectively ordered with respect to their energies, or the contribution they make to the reconstructed object. A significance layer is generated

by choosing a bit-position and outputting the value of that bit-position (1 or 0) for all the samples for which that bit-position is defined. Hence, in Figure 6, 'R' is the most significant bit of a component's coefficients and 'r' is the bit-significance at which the encoder stops refinement.

5

An embedded representation is thus obtained by starting with the most-significant bits of the coefficients, and sending successively lower-order bits, until a desired quality is reached. From the point of view of a particular sample, the sample is first approximated by having its most-significant bit set, and successively approximated to greater precision as lower-significance layers become available. The process of increasing the precision of a sample is called refinement.

10

An encoding of source material can be structured in different ways, and use different combinations of layers, while delivering exactly the same information. For example, a representation of an entire video at low resolution may come before other layers, each of which causes a step in the quality of the entire video. Alternatively all the information for a particular frame may appear before any information for the next in which case reading the representation sequentially will produce one frame after the other, each delivered to as high a quality as possible. However, given signalling to delimit individual layers all structures are equivalent since one can be processed into another.

20

The primary use of layer signalling is to allow a hierarchical, structured and layered representation to be compiled into a plurality of sequential, unstructured and embedded representations for use by multiple clients. Figures 7 to 11 are examples of such representations which may be obtained from a single layered object. In figure 7 all the layer information has been transmitted and used exactly to reconstitute the original. In figures 8 and 9 only selected scales have been used and result in an image with reduced spatial resolution and frequency content (a reduced-scale image). In figure 10 information from all scales has been transmitted so full spatial resolution is obtained, but the significance layers have been truncated so the image is distorted because lower-energy wavelet coefficients have not been transmitted to their full precision. In figure 11 a truncated significance layer has been transmitted for all scales, except those which encode horizontal details, which accordingly appear in the reconstructed image to the original fidelity.

25

30

35

Appendix 1

Some Definitions of Terms, as used in the context of the illustrated embodiments

Scale

- 5 A function can be analysed into a set of components with different time/space/frequency content. These components are called scales, and the process of analysis into scales is called multi-scale decomposition. The analysis is performed by a waveform of limited duration and zero integral, called a wavelet.
- 10 Components that are highly localised in space or time but have a ill-defined frequency spectrum are small-scale and capture fine detail. Components that are spread through space but have a precisely-defined frequency spectrum are large-scale and capture general trends.

Embedding

- 15 Encoding a source object in such a way that the generation and/or reception of the encoded bitstream can be terminated at any point, the resulting representation being the 'best' attainable, according to some criteria set by the encoder.

Layer

- 20 A layer is structure with a particular format: a media stream may be organised by an encoder into a tree-structured hierarchy of Layers, where a Layer represents a node in the tree and it's associated subtree. The subtree consists, either of further Layers, or of a Simple Layer which consists of an embedded bitstream.

Base Layer

- 25 A layer which is sent first to a client to initialise a representation of a media object at that client, to which refinement layers are added. The Base Layer usually carries 'coarse-scale' or 'trend' information which allows a general impression of the media to be obtained, although any layer can be used as a Base Layer.

30

Refinement Layer

- A Layer which is sent to a client subsequent to a Base Layer and which reduces the distortion of the representation of a media object at that client. The Refinement Layer usually carries 'fine-scale' or 'detail' information which enhances the representation existing
- 35 at the client.

Simple Layer

A layer which contains embedded refinement information only, i.e., there are no further sub-layers.

5

Layer Embedding

The process of writing Layers to a bitstream in such a way that those layers carrying the 'best' information, according to some criteria which may be set by the encoder, appear early in the sequence.

10

Significance Layer

A layer where all the refinement information refers to a particular bit-position for all the coefficients undergoing refinement.

Scale Layer

A layer for which all the refinement information is for a particular scale.

Region Layer

A layer for which all the refinement information is for a particular connected region in a space or time-varying function. This enables different regions to be studied in different ways, with information being concentrated in different combinations of Layers according to the requirements of the bitstream. For example, a 'hi-fidelity' window can be defined over areas of particular interest.

Significance-Scale Layering.

A layering scheme whereby Significance Layers are embedded within Scale Layers. Such a scheme is important because it allows a trade-off to be made between bit-rate, spatial resolution and distortion in the bitstream transmitted to the client.

Distortion

The distortion of an image can be measured in terms of its Peak Signal-to-Noise Ratio (PSNR) where $PSNR = 10\log(255^2/MSE)$ dB, and MSE is the image's Mean Squared Error.

35

Claims

1. A method of delivering a media file to a client on a network in which a server derives the media file from a source file on the server and delivers the media file to the client,
5 the derivation and/or delivery being in compliance with a set of download parameters, the download parameters defining one or more download and/or media variables.
2. The method of Claim 1 wherein the set of download parameters is stored on the
10 server.
3. The method of Claim 1 wherein the download parameters are transmitted to the
server by the client before or during the delivery of the media file.
4. The method of Claim 1 wherein the download parameters are a combination of
15 values stored on the server and values transmitted to the server by the client before or during the delivery of the media file.
5. The method of Claim 1 wherein the download parameters define one or more of
the following media parameters for an image or sequence of images:
20 (i) the spatial resolution of the image or images;
(ii) the quality of the image or images;
(iii) the number of displayable frames;
(iv) the preferred order in which data is to be transmitted;
(v) the selection of colour components within one or more frames;
25 (vi) a sub-set of frames to be delivered; and
(vii) a sub-region within one or more frames;
6. The method of Claim 1 wherein the download parameters define one or more of
the following media parameters for audio:
30 (i) the quality of the audio;
(ii) the preferred order in which data is to be transmitted;
(iii) the number of audio channels to be transmitted; and
(iv) selection of monophonic, stereophonic or quadraphonic audio.

7. The method of Claim 1 wherein the download parameters define one or more of the following:

- (i) the rate at which data is to be transmitted to the client;
- (ii) the preferred order in which data is to be transmitted;
- 5 (iii) the set of data for the source file that is already stored on the client and does not need to be transmitted; and
- (iv) the maximum amount of data to be transmitted to the client;

8. The method of Claim 1 in which the derivation of the media file is adapted to take
10 account of one or more of the following:

- (i) the data size of the original of the source file
- (ii) the bandwidth available to a client.
- (iii) the current or predicted loading of the server

15

9. The method of Claim 1 in which any one or more of the download parameters may be altered by the client after the transfer of the image files has completed and any extra data required to satisfy the revised download parameters are then transferred from the server.

20 10. The method of Claim 1 in which the media file is generated using a wavelet transform.

11. The method of Claim 10 in which the output of the wavelet transform is compressed using SPIHT compression.

25

12. The method of Claim 1 in which the media file is derived such that the quality of the media file at the client progressively increases as additional data is downloaded to the client.

13. The method of Claim 1 in which the media file is derived such that the number of
30 displayable frames in a sequence progressively increase as additional data is downloaded to the client.

14. The method of Claim 1 in which some or all of the media file is retained on the client and subsequent use of the source file by the client at a different resolution, quality or
35 frame rate utilises the retained data together with any additional data that is required.

15. The method of Claim 1 in which the data in the media file is structured by an encoder as a bitstream including several discrete bitstream layers, in which layer signalling information which identifies individual layers is inserted by the encoder, the layer signalling information enabling each client to be sent only those layers which satisfy the download parameters specified by that client.
16. A media file which is deliverable using any of the methods defined above.
17. A computer program which when running on a client enables the client to receive and playback a media file delivered using any of the above methods.
18. A computer program which when running on a server or encoder enables the server or encoder to perform any of the above methods.
19. A server programmed to deliver to a client on a network a media file deriving from a source file stored on the server, wherein the server is programmed to derive the media file from a source file on the server and deliver the media file to the client, the derivation and/or delivery being in compliance with a set of download parameters, the download parameters defining one or more download and/or media variables.

Networked delivery of profiled media files to clients

Abstract

- 5 A server transfers to a networked client a media file in compliance with download parameters previously transmitted to the server from the client, the download parameters defining one or more download and/or image variables. Because the download and/or image variables are controlled by each client, a video or audio file can be played back at different clients at different data rates, resolutions and quality levels

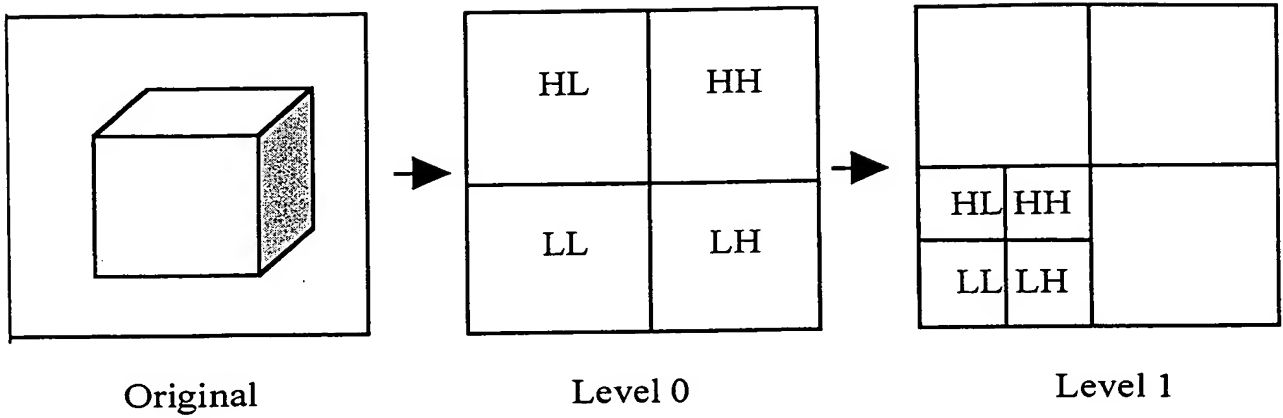


Figure 1

Figure 2

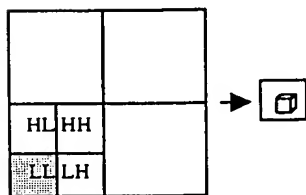


Fig 2A

Reconstruction
using LL
Subband of level 1
(lowest-resolution)

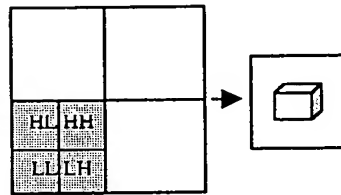


Fig 2B

Reconstruction
using LL
Subband of level 0
(medium
resolution)

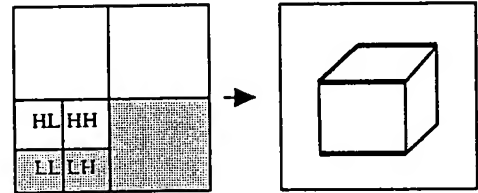
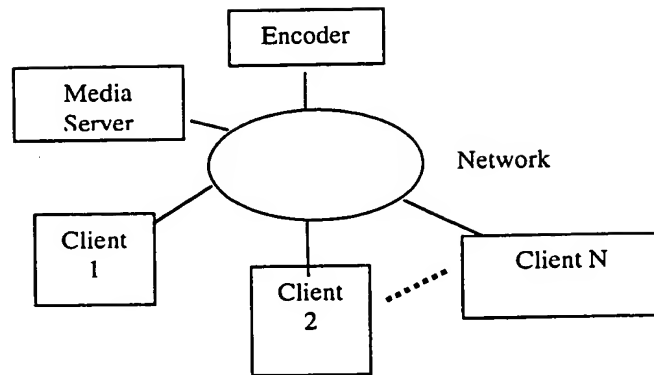
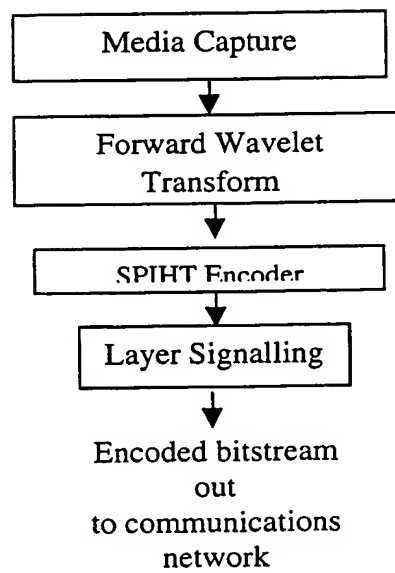


Fig 2C

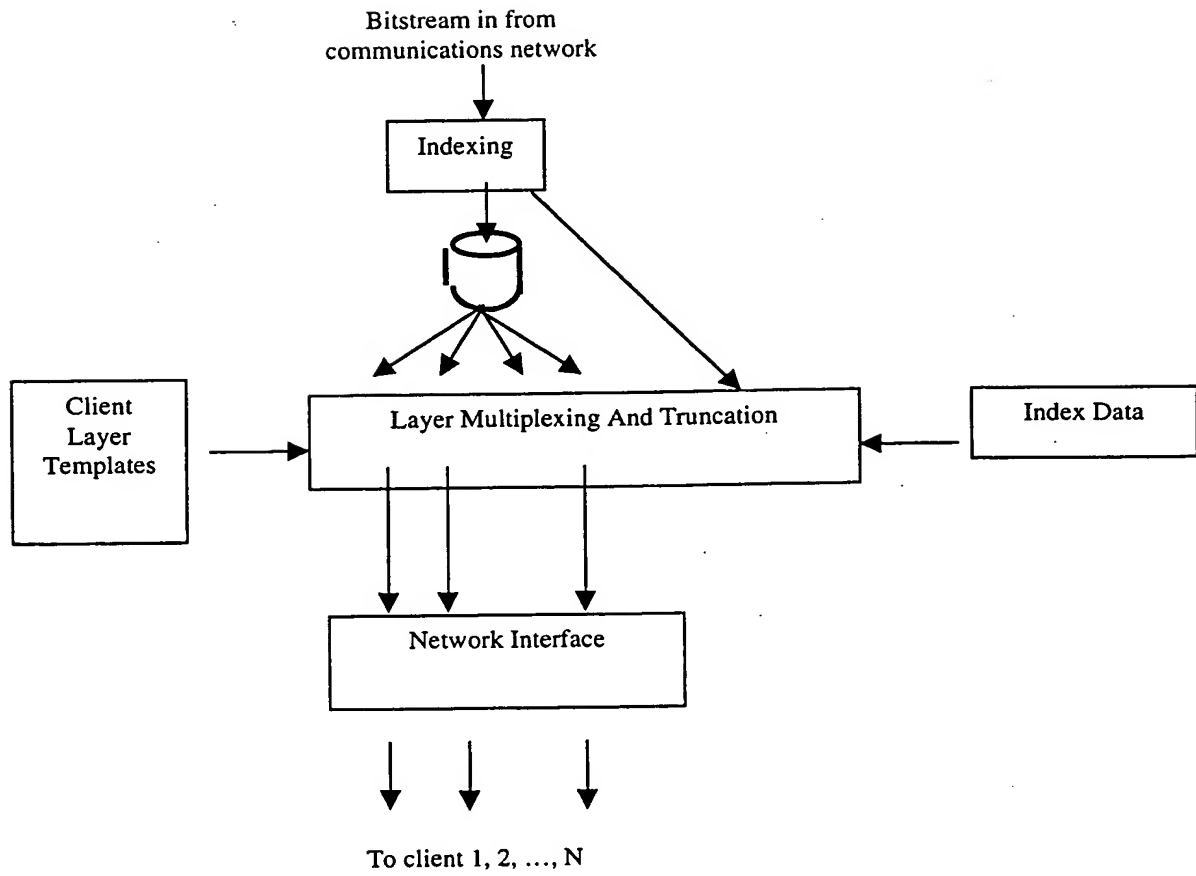
Reconstruction using LL & LH,
subband
of level 1
And LH subband of level 0
(Original resolution,
horizontal features at all scales
emphasised)

THIS PAGE BLANK (USPTO)

**Figure 3****Figure 4**

THIS PAGE BLANK (USPTO)

Figure 5



THIS PAGE BLANK (USPTO)

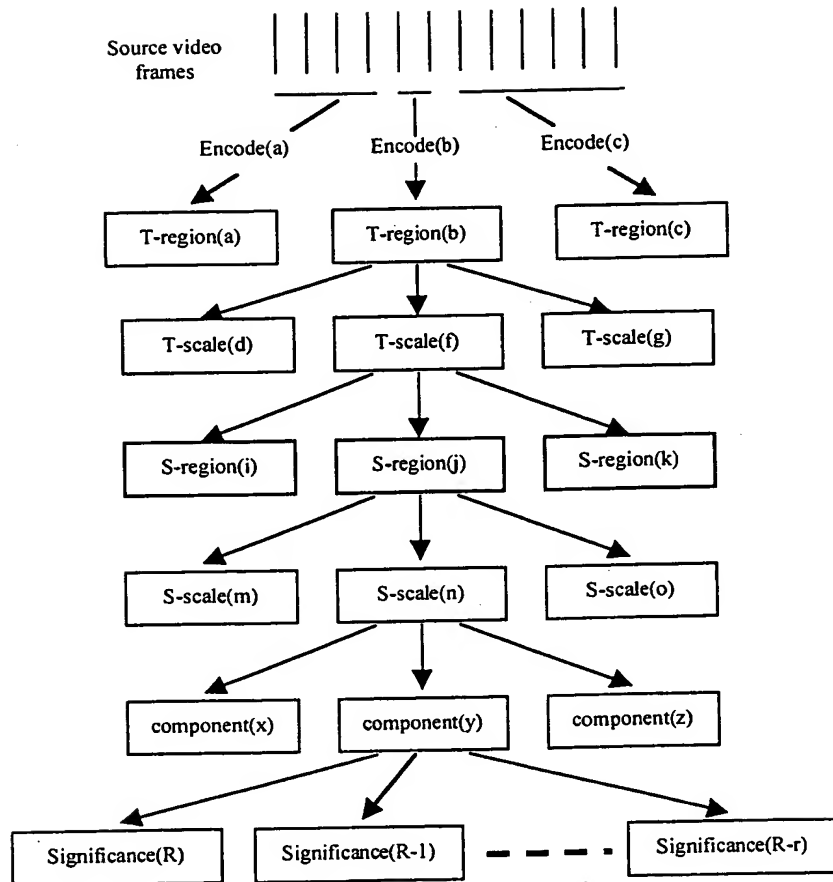
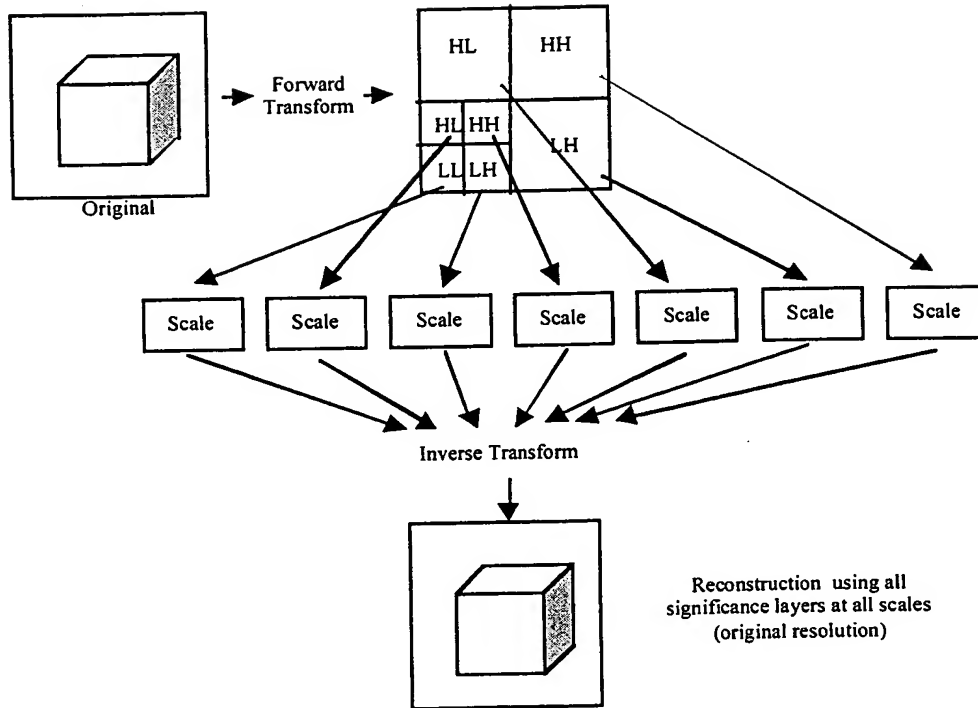
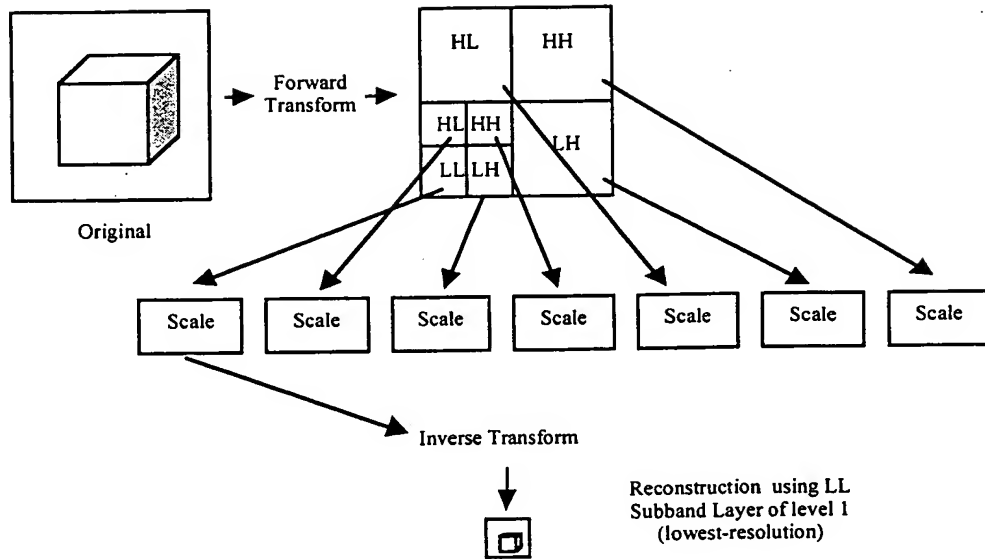


Figure 6

THIS PAGE BLANK (USPTO)

**Figure 7**

THIS PAGE BLANK (USPTO)

**Figure 8**

THIS PAGE BLANK (USPTO)

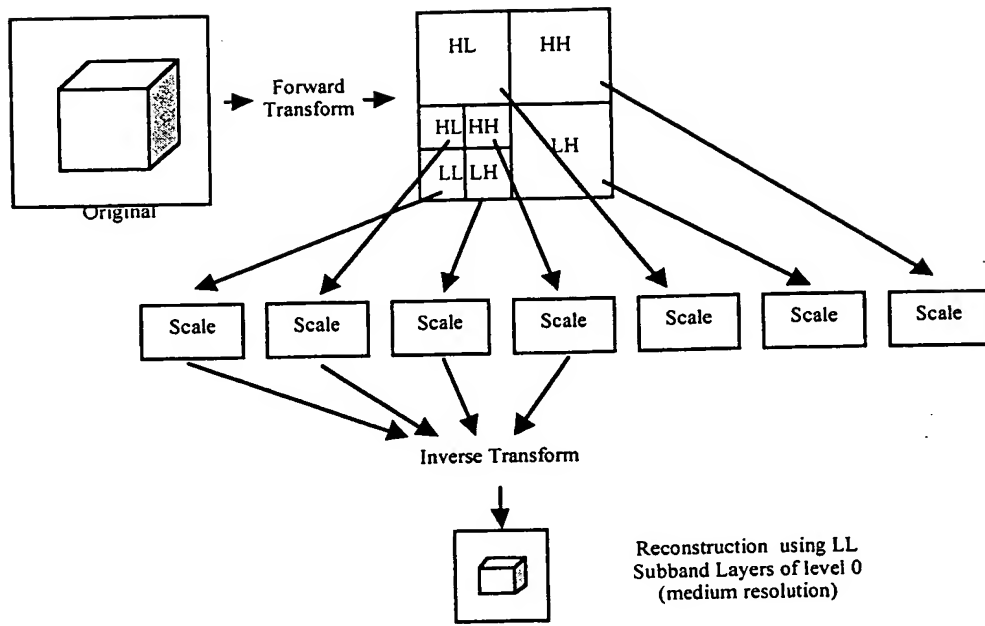


Figure 9

THIS PAGE BLANK (USPTO)

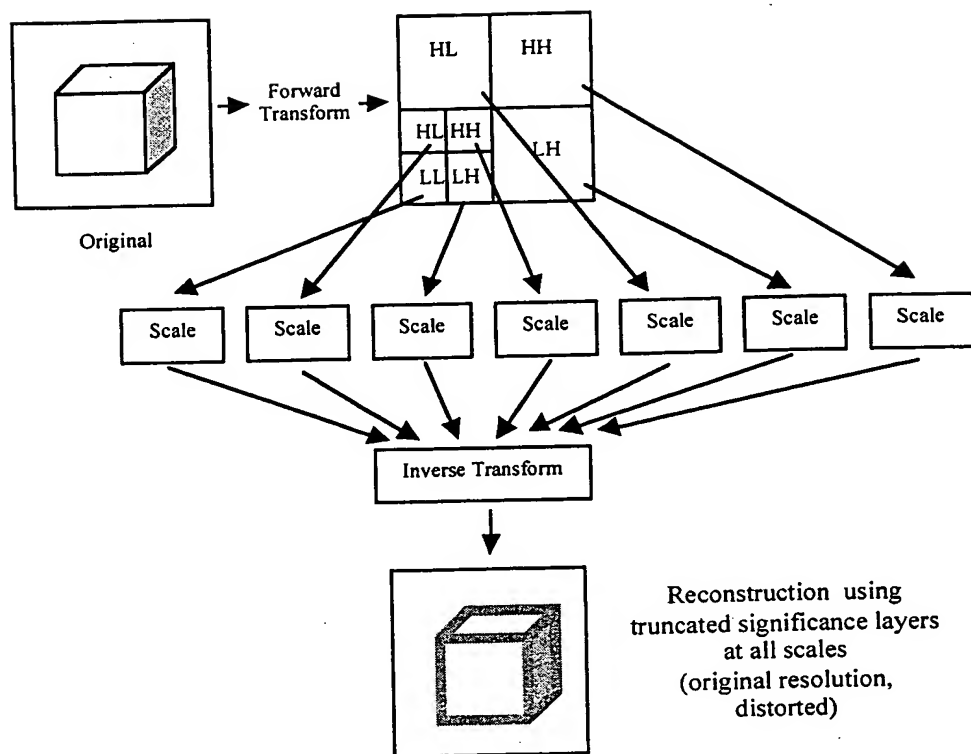


Figure 10

THIS PAGE BLANK (USPTO)

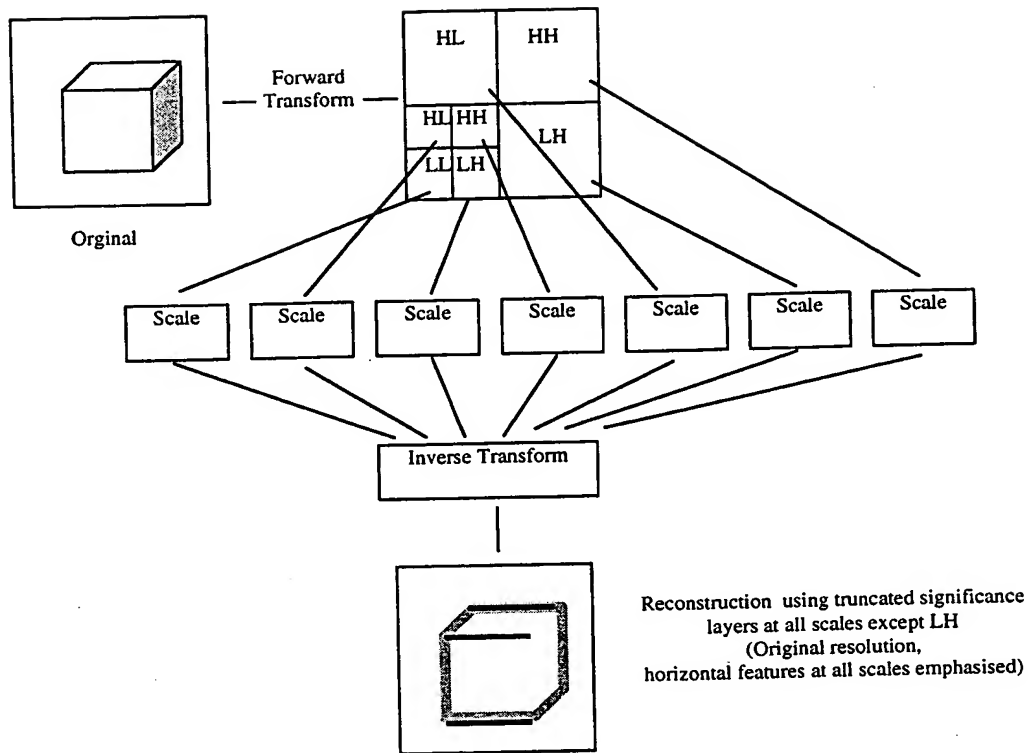


Figure 11

PCT / GB00 / 016 10

24.7.00

origin

THIS PAGE BLANK (USPTO)